

What Makes a Link Successful on Wikipedia?

Dimitar Dimitrov*

GESIS – Leibniz Institute for the Social Sciences
dimitar.dimitrov@gesis.org

Philipp Singer*

GESIS – Leibniz Institute for the Social Sciences
& University of Koblenz-Landau
philipp.singer@gesis.org

Florian Lemmerich

GESIS – Leibniz Institute for the Social Sciences
& University of Koblenz-Landau
florian.lemmerich@gesis.org

Markus Strohmaier

GESIS – Leibniz Institute for the Social Sciences
& University of Koblenz-Landau
markus.strohmaier@gesis.org

ABSTRACT

While a plethora of hypertext links exist on the Web, only a small amount of them are regularly clicked. Starting from this observation, we set out to study large-scale click data from Wikipedia in order to understand *what makes a link successful*. We systematically analyze effects of link properties on the popularity of links. By utilizing mixed-effects hurdle models supplemented with descriptive insights, we find evidence of user preference towards links leading to the periphery of the network, towards links leading to semantically similar articles, and towards links in the top and left-side of the screen. We integrate these findings as Bayesian priors into a navigational Markov chain model and by doing so successfully improve the model fits. We further adapt and improve the well-known classic PageRank algorithm that assumes random navigation by accounting for observed navigational preferences of users in a weighted variation. This work facilitates understanding navigational click behavior and thus can contribute to improving link structures and algorithms utilizing these structures.

Keywords: Human Click Behavior; Navigation; Wikipedia

1. INTRODUCTION

Even though links are omnipresent on the Web, only a minority of them get regularly clicked by humans. For example, on Wikipedia only around 4% of all existing links are clicked by visitors more frequently than 10 times within a month (cf. Section 3). This phenomenon demonstrates the importance of a deeper understanding of human navigation behavior in order to provide efficient and effective navigational support for users. While a variety of navigational regularities, patterns and strategies have been identified in previous work [26, 31, 36, 42], our research community still lacks a systematic understanding of *factors that make a link successful in information networks*. Towards that end, we turn our attention to Wikipedia, one of the most popular information sources on today’s Web.

Problem and objectives. In particular, we are interested in understanding *the relationship between link properties and link popularity* as measured by *large-scale transitional click data*. Understanding what link features effect user click behavior can have implications for how we place links on Wikipedia by e.g., better mirroring human

navigational choices, but can also have consequences for fundamental Web algorithms such as Google’s PageRank that assume random navigation casting all links as equal and not account for potential preferences towards specific links.

Materials, approach and methods. We utilize openly available large-scale datasets about the English Wikipedia: the network of internal links within Wikipedia articles and navigational data capturing user transitions between articles (cf. Section 2). Motivated by previous work, we study three types of features that we compute from the link network and full texts of the articles: (i) *structural features* (e.g., humans might prefer to navigate to central nodes in the information network) [42], (ii) *semantic features* (e.g., humans might prefer to navigate between semantically similar articles) [38] and (iii) *visual features* (e.g., humans might prefer links on the top of the screen) [13]. For example, in Figure 1, given the Wikipedia article “Manchester” and the out-going links to the articles “England” and “Association Football”, we are interested in understanding which link features (e.g., the visual position of the link) can better predict the actual transitions (popularity) of a link.

Methodologically, we start by gaining descriptive insights into the success of links before we statistically model corresponding effects using mixed-effects hurdle models [18, 32] allowing us to account for the heterogeneity in the data (e.g., links in the article “USA” might in general be more frequently used compared to links in the article “Manchester”). Subsequently, we integrate obtained insights into existing stochastic models of human navigation. To that end, we utilize the first-order Markov chain model—probably the most popular model for this task—in two separate analyses. First, we craft hypotheses about human navigation based on our insights and integrate these into a Bayesian inference process as priors; cf. [36]. Based on Bayesian model selection, this allows us to identify whether given hypotheses can improve the Markov chain model fit compared to an uninformed model; additionally, we can obtain a ranking of hypotheses allowing us to confirm our statistical results within a coherent research approach. Second, we aim at further utilizing these hypotheses to advance the classic well-known PageRank algorithm in a weighted variation relaxing the basic assumption of an uninformed random surfer; we evaluate the ranking of articles against actual page views.

Contributions and findings. The main contributions of this paper are three-fold: (i) We provide empirical evidence that Wikipedia users have a preference of choosing links leading to the periphery of underlying topological link network, that they prefer links leading to semantically similar articles, and that links positioned at the top

*Both authors contributed equally to this work.

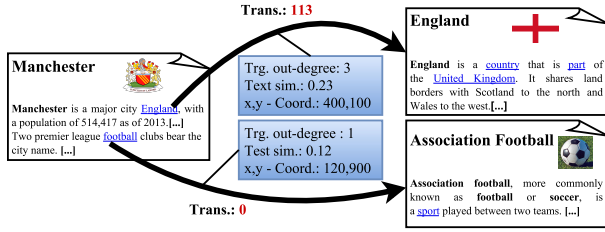


Figure 1: **Example.** Wikipedia pages are connected by links for which we can compute a variety of features based on *network features* (e.g., target article out-degree), *semantic features* (e.g., text similarity between source and target articles), or *visual properties* (e.g., position of the link on the screen). This work aims at understanding what makes a link successful—i.e., which link properties best explain observed numbers of user transitions (shown on edges).

and left-side of an article have a higher likelihood of being used. (ii) We integrate this evidence with first-order Markov chain models demonstrating an improvement of respective model fits by incorporating the assumed behavior of human navigation into the inference process (Section 5.1). (iii) Finally, we demonstrate the utility of these findings by adapting the well-known PageRank algorithm to better reflect human navigation behavior. Our enhancements lead to a significant improvement over the uninformed baseline algorithm that assumes random navigation by evaluating obtained ranking against actual article page views on Wikipedia (Section 5.2). The methodological framework provided in this work can also easily be applied to other transitional click data apart from Wikipedia.

2. DESCRIPTION OF DATA

This section describes the utilized datasets and extracted features. We make an implementation of the data extraction process as well as a sample of the data publicly available online.¹

2.1 Wikipedia link and transition data

Wikipedia data. In this work, we focus on all articles contained in the main namespace of the English Wikipedia as extracted from the public XML dump from March, 2015.² To obtain authentically rendered pages, we retrieved the corresponding static HTML pages by using Wikimedia’s public API. In contrast to using readily available link dumps, this allowed also for considering links that are indirectly included in a page, e.g., by templates. A tiny part ($< 0.01\%$) of articles could not be retrieved and had to be excluded from the analysis. With this data, we created the Wikipedia link network D_{wiki} using articles as nodes and unique links as directed edges; D_{wiki} contains ~ 4.8 million articles connected by ~ 340 million distinct links.

Transition data. For measuring actual usage of links, we utilize openly available transition data from Wikipedia from February 2015 [45]. It contains aggregated page requests extracted from the server log for the English desktop version of Wikipedia in the form of (*referrer*, *resource*) pairs, i.e., transitions, and their respective transition counts. The data has already been pre-processed to filter bots and web crawlers and transitions occurring less than 10 times, see [45] for details. Since in this paper we are only interested in studying *internal* navigation on Wikipedia, we excluded article requests from outside of Wikipedia (e.g., users arriving directly from search engines) and focus only on transitions corresponding to an internal link in our network D_{wiki} . These amount for about 1 billion

or 31% of all page requests and cover ~ 13.6 million distinct links between ~ 1.4 million source and ~ 2.1 million target articles. With this information we can weight edges in D_{wiki} by their transition counts to obtain an edge-weighted network D_{trans_w} . Additionally, we denote the sub-network of D_{wiki} that contains only edges (links) that we observe in the transition data—i.e., those links having a weight larger than 0 in D_{trans_w} —as D_{trans} .

2.2 Link features

For studying link success, we focus on three types of link features: network features, semantic similarity features, and visual features.

Network features. Here, we capture the centrality of the source (*src*) and target articles (*trg*) of a specific link in the Wikipedia link network D_{wiki} considering the *in-degree*, the *out-degree*, the overall *degree*, the *pagerank* [30] and the *k-core* [2] measure.

Semantic similarity features. Here, we detect the semantic similarity between the source and target article of a specific link. First, we computed a *text similarity* score by utilizing a sub-linear scaled tf-idf weighted vector space model capturing word tokens of Wikipedia articles [36]. For dimensionality reduction, we used sparse random projection retaining the pairwise distance between concepts with some small error. The text similarity between two pages was then determined by the cosine similarity of corresponding vectors. We also computed a *topic similarity* score with the cosine similarity between Wikipedia categories assigned to source and target.

Visual features. The third group of features is based on the placement of links within the source article, cf. also [13]. For that purpose, we rendered the HTML of each articles on a screen for a resolution of 1920×1080 , a common resolution for desktop users. This enables deriving the exact screen position of links in terms of their *x/y-coordinates*. Based on visual position, CSS classes, and HTML tags, we additionally assigned each link to one of the following visual regions: (1) *lead*: all links in the first section of the article excluding infobox, (2) *body*: all other links in the main text, (3) *left-body*: links in the body that are displayed in the 10% most left part of the screen, (4) *right-body*: links in the remaining right part of the body, (5) *infobox*: all links in the infobox, (6) *navbox*: all links in a table in the last section of the article placed by editors for facilitating navigation. As links can occur multiple times on an article, we derived the visual position of a link based on its first occurrence on the screen. This goes hand-in-hand with our empirical results elaborated in Section 4.2. A more detailed discussion about this facet of our data is provided in Section 7.

3. FOCUS OF ATTENTION

We start our analysis by investigating the focus of attention with respect to link usage in Wikipedia, i.e., we study how strongly user transitions are concentrated on the few most visited links.

Distribution of transition counts. As a first key statistic, we observe that of ~ 340 million links in D_{wiki} only about 13.6 million links (or about 4% of all links) are regularly used, i.e., are visited at least 10 times within a month. However, even when considering only these 4% regularly used links, we can observe that users focus heavily on a comparatively small set of links: about 50% of all transitions stem from only about 600,000 links. A frequency plot of the distribution, see Figure 2, reveals a long-tailed distribution, where most of the links receive very low amounts of transitions.

Out-degree distribution. A similar effect can be observed at the level of individual articles. We compare the out-degree distributions of nodes in D_{wiki} and D_{trans} , i.e., we compare the number of available links on articles in D_{wiki} with the number of regularly

¹<https://github.com/linksuccess/linksuccess>

²<https://archive.org/details/enwiki-20150304>

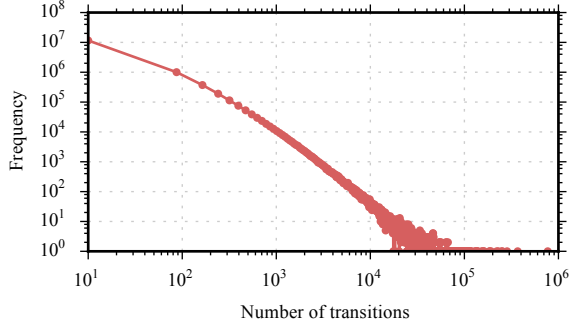


Figure 2: **Distribution of transitions counts.** This figure shows the frequency distribution of link transition counts. The long tail distribution shows that majority of links have low popularity.

used links in D_{trans} . For a fair comparison, we consider only the out-degrees of nodes that occur in both networks. The results in 3 reveal clear differences. The out-degree distribution of D_{wiki} reaches a maximum frequency for articles having eleven out-links indicating a decent amount of links to choose from. However, the out-degree distribution of D_{trans} clearly reveals that for most articles, users only visit a few articles regularly with the maximum frequency of only a single link. Formal comparison of both distributions with respect to (truncated) power-law, exponential, and log-normal distributions by their goodness of fit reveals that for D_{trans} the out-degree distribution resembles a truncated power-law most closely, while for D_{wiki} a log-normal distribution is preferred.

Gini coefficient of transition counts. The previous analyses have suggested a discrepancy between the link network D_{wiki} and actually transitioned links in D_{trans} and $D_{trans,w}$. The *Gini coefficient* is a measure of statistical dispersion allowing us to quantify the inequality of link popularity within articles. We calculate the Gini coefficient for each article based on a vector containing the transition counts for all links (including zeros) in this article from $D_{trans,w}$. A Gini coefficient of 0 indicates complete equality, while a result of 1 would mean that only a single link gets the whole attention. We visualize the histogram of all coefficients for each article in Figure 4; results indicate that there is a clear inequality of link popularity within articles, i.e., attention focuses on just a few links.

Summary. In this section, we have presented empirical evidence that on Wikipedia user attention focuses on a small set of links while navigating. This hints towards a clear *preference towards a few links* and demonstrates the importance of finding indicators of link success, on which we focus in the remainder of this paper.

4. WHAT MAKES A LINK SUCCESSFUL?

Motivated by the discrepancy between link presence and link usage, we now study link success. That is, as illustrated in Figure 1, given a Wikipedia article, we aim at identifying effects of link features (network features, semantic features, or visual features) on counts of transitions to the respective target article. This allows us to answer questions such as: “Are links leading to peripheral articles more popular and thus, more successful?” We propose a methodology in Section 4.1 and discuss results in Section 4.2.

4.1 Methodology

For studying effects of link features on the success of links, we employ descriptive statistics and *hurdle mixed-effects regression models*; we outline both approaches next.

Descriptive analysis. We start our analysis with a descriptive

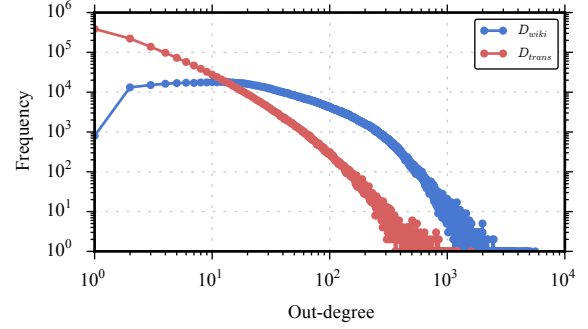


Figure 3: **Out-degree distributions.** This figure compares the out-degree distributions of the Wikipedia link network D_{wiki} (blue) and the transition network D_{trans} (red) for nodes contained in both networks. We can observe that only a few links are visited regularly on articles (peak at out-degree of one in D_{trans}) while more links are available (peak at out-degree of eleven in D_{wiki}).

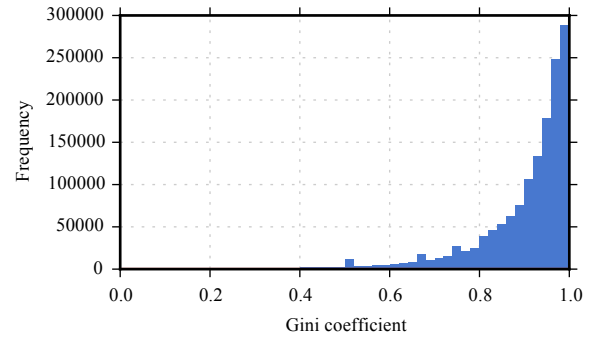


Figure 4: **Distribution of Gini coefficients.** The histogram of Gini coefficients calculated on article vectors of link transition counts reveals clear inequality of link popularity within articles.

overview on the effects of our features on transition counts. For an intuitive picture, we investigate boolean expressions on our features and determine for feature the number of links, the number of transitions, and the average number of transitions per link. We focus on easy-to-interpret expressions: for network features, we compare the centrality of the target node with the centrality of the source node. For semantic similarity features, we compare the similarity between source and target article of the link with the median of the respective values for all links within the same source article. Finally, for the visual features, we divide the screen position in three areas for each dimension, that is the left, the middle, and the right third of the screen, and links in the top-half, the bottom-half and links that are only visible after vertical scrolling (assuming a resolution of 1920×1080) respectively. Furthermore, we use the position of the link in a specific part of the article such as the lead or the infobox.

Mixed-effects hurdle model. The methods elaborated in this section aim at statistically modeling the effects of all link features on link success given a certain start node. Our setting of interest (cf. Figure 1) implies that individual links represent data points with the transition counts being the outcome to predict, and the respective link features the effects at interest. Yet, links in the dataset are not independent of each other, but rather nested within Wikipedia articles. Heterogeneous effects might be present as e.g., apart from generally different amounts of total transitions, the article of “Manchester” might pose different navigational behavior than the article of “USA”. Pooling all links into a single dataset for ordinary regression can lead to biased results; consequently we resort to *mixed-effects models*

(also known as multilevel models or hierarchical models), allowing us to account for the nested structure—specifically the resulting random variations—between links (level 1) and Wikipedia pages (level 2). We define our mixed-effects model following [17]; for a broader introduction we point to [32]:

$$Y_{i,j} = \gamma_{0,0} + \gamma_{1,0}X_{1,i,j} + u_{0,j} + u_{1,j}X_{1,i,j} + e_{i,j} \quad (1)$$

Here, $Y_{i,j}$ refers to the transition count of a link i (level 1) on a Wikipedia page j (level 2); $\gamma_{0,0}$ refers to the overall intercept and $\gamma_{1,0}$ to the overall regression coefficient for the level 1 predictor $X_{1,i,j}$, e.g., the degree of the target node. These are the fixed effects. The remainder refers to the random effects: $u_{0,j}$ is the error component of the intercept allowing the group intercepts to differ, $u_{1,j}$ refers to the error component of the slope allowing the group slopes to differ, and $e_{i,j}$ is the classic random error of predictions.

To study the effects of network, semantic similarity and visual features, we model each feature separately. As shown in Section 3, only 4% of all links have been used at least 10 times, resulting in a large amount of zeros in the data outcomes. To account for that, we employ *hurdle models* [18] in a two-step approach. (i) First, we model the binary data in D_{trans} ; i.e., the link values are set to 1 iff at least 10 transitions for these links have been observed, and to 0 otherwise. We then model this data with a mixed effect binomial logistic regression and make inference on the fit. (ii) In a second step, we only take the data from D_{trans_w} capturing those links that have been clicked and their respective transition counts. As our outcome variable is generated by count data, basic assumptions for *linear* mixed-effects models are violated. Consequently, we resort to zero-truncated negative binomial mixed-effect models that do not allow the outcome to be zero (as all outcome counts are at least 10). The choice of model also enables modeling the presence of overdispersion in the data. For comparison, we also fitted Poisson mixed-effects models with an observation-level random effect that allows for overdispersion [14] but could not find core differences in inference. Thus, we overall build two models for each feature: one to model *if* a link has been used at least 10 times in a month, and a second to model the transition *counts* for these regularly used links.

We fit our models using the “Template Model Builder” framework [25] that evaluates and maximizes the Laplace approximation of the marginal likelihood utilizing automatic differentiation of the joint likelihood; we use the R package “glmmTMB”³. For improving convergence, we scale non-binary features by subtracting the mean and dividing by their standard deviation. For judging significance of individual effects, we follow an incremental model comparison approach [3]. That is, we compare the model including the effect of interest with a restricted model that neglects the effect using a *likelihood ratio chi-square test* (AIC and BIC agree).

4.2 Results

For computational reasons, we applied the methodology to a random sample of 10,000 articles containing at least one out-going link in the transition data D_{trans} considering all links and their features. The sample contains 1,028,704 links with 6,686,581 transitions. We made sure that the sample captures similar power-law degree distributions as the overall dataset (cf. Section 3); multiple sampling iterations did not change the results. For reproducibility reasons, we make the full sample available.¹ We present the descriptive results in 1 and the mixed-effects model results in 2. For the statistical models, we only report fixed effects at interest but make the full results available in additional material.¹ We discuss the results for our three major feature groups next.

Network features. The results in 1 highlight that there appears to be a preference of users to choose links leading to target Wikipedia articles that have a smaller in-degree, out-degree and degree compared to the source article. This is specifically imminent when controlling for the presence of links as the last column indicates. Similar behavior can be seen for the k-core and pagerank of nodes. Our mixed-effects model results shown in 2 statistically confirm this observation. The coefficients for both parts of the hurdle model (binomial and zero-truncated negative binomial models) show negative signs for all network features at hand meaning that the higher those features are (i.e., the more central the target nodes are) the less likely it is that those links are clicked. Interestingly, the effect sizes and model fits (LRT) are quite similar for all features except the out-degree which can be explained by the fact that in general the out-degree is not a good indicator regarding the centrality of a node in the network as it just depends on the number of links present on a given page. Overall, the results suggest a preference for users to navigate to peripheral nodes in the network in contrast to navigating towards the core of the network.

Semantic similarity features. The results in 1 and 2 indicate that links connecting two semantically related Wikipedia articles—both for text and topic similarity—are more likely to be transitioned compared to those links that indicate less semantic similarity. This is imminent from the higher ratio of transitions to high similarity target articles, and specifically, by the positive mixed-effects model coefficients for the hurdle models at hand. Overall, the results suggest a preference towards navigating semantically similar articles supporting previous research [36, 38, 42].

Table 1: **Descriptive results.** This table provides a descriptive overview on link success; it shows for each boolean feature the number of links, the overall number of transitions these links accumulated, and the average transition count per link. Results indicate that links leading to the periphery of the network, to semantically similar articles, and links positioned in the lead and left area of an article accumulate relatively more transitions than opposite links.

Feature	# Links	# Transitions	Mean Trans. per Link
trg_degree < src_degree	462,147	3,944,747	8.54
trg_degree ≥ src_degree	566,557	2,741,834	4.84
trg_in_degree < src_in_degree	412,653	3,831,100	9.28
trg_in_degree ≥ src_in_degree	616,051	2,855,481	4.64
trg_out_degree < src_out_degree	545,980	3,912,604	7.17
trg_out_degree ≥ src_out_degree	482,724	2,773,977	5.75
trg_kcore < src_kcore	283,668	3,477,828	12.26
trg_kcore ≥ src_kcore	745,036	3,208,753	4.31
trg_page_rank < src_page_rank	414,124	3,833,596	9.26
trg_page_rank ≥ src_page_rank	614,580	2,852,985	4.64
text_sim > median of article	511,871	4,335,445	8.47
text_sim ≤ median of article	516,833	2,351,136	4.55
topic_sim > median of article	392,955	3,246,766	8.26
topic_sim ≤ median of article	635,749	3,439,815	5.41
position = lead	93,546	2,489,342	26.61
position = body	373,407	3,442,017	9.22
position = left-body	64,320	934,766	14.53
position = right-body	309,087	2,507,251	8.11
position = navbox	502,079	99,498	0.2
position = infobox	59,672	655,724	10.99
screen_x_coord: left third	211,549	2,358,899	11.15
screen_x_coord: middle third	291,260	1,422,183	4.88
screen_x_coord: right third	525,847	2,905,499	5.53
screen_y_coord: top half	221,980	3,991,912	17.98
screen_y_coord: bottom half	174,500	998,287	5.72
screen_y_coord: scroll needed	632,176	1,696,382	2.68
Overall	1,028,704	6,686,581	6.5

³<https://github.com/glmmTMB/glmmTMB>

Table 2: **Mixed-effects hurdle modeling results.** This table presents the results for the mixed-effects hurdle models based on a basic model specification of: $\text{transition}_{i,j} = \gamma_{0,0} + \gamma_{1,0}\text{feature}_{1,i,j} + u_{0,j} + u_{1,j} \text{feature}_{1,i,j} + e_{i,j}$. For each feature (first column) and a given transformation (second column), results for both parts of the hurdle model are reported. The binomial models the binary outcome whether a link has been used at all, and the zero-truncated negative-binomial models the transition counts of regularly used links. For each model, we report the fixed effect coefficient of the fitted model, the likelihood ratio test statistic compared to a reduced nested model without the fixed effect, and the significance according to χ^2 -test. Overall, the results confirm the descriptive results of Table 1 suggesting a preference of links leading to the periphery of the network, links targeting semantically similar articles, and links being positioned at the beginning and left-side of articles.

Feature	Transformation	Binomial Model			Zero-truncated NB Model		
		Fixed effect	LRT	Pr(>Chi)	Fixed effect	LRT	Pr(>Chi)
trg_degree	scale	-9.0704	9441	< 1.0 e-300 ***	-0.1353	174	7.98 e-40 ***
trg_in_degree	scale	-9.2925	9456	< 1.0 e-300 ***	-0.1341	185	4.2 e-42 ***
trg_out_degree	scale	-0.6425	4052	< 1.0 e-300 ***	-0.0375	32	1.6 e-08 ***
trg_kcore	scale	-9.4873	9470	< 1.0 e-300 ***	-0.1296	180	4.5 e-41 ***
trg_pagerank	scale	-8.8103	9384	< 1.0 e-300 ***	-0.1457	196	1.7 e-44 ***
text_sim	scale	0.2944	617	3.4 e-136 ***	0.238	1179	1.9 e-258 ***
topic_sim	scale	0.2052	482	8.6 e-107 ***	0.1212	401	3.7 e-89 ***
position = lead	none	1.9682	6146	< 1.0 e-300 ***	0.2858	498	2.7 e-110 ***
position = body	none	0.2351	93	5.9 e-22 ***	-0.4581	1478	< 1.0 e-300 ***
position = left-body	none	1.106	739	9.4 e-163 ***	-0.1671	123	1.7 e-28 ***
position = right-body	none	-0.3631	231	4.4 e-52 ***	-0.4389	1236	1.0 e-270 ***
position = infobox	none	-0.1955	25	7.2 e-07 ***	0.1365	37	1.4 e-09 ***
position = navbox	none	-6.3237	9226	< 1.0 e-300 ***	-0.9963	688	1.5 e-151 ***
screen_x_coord	scale	-0.2974	1081	4.4 e-237 ***	0.0225	16	6.0 e-05 ***
screen_y_coord	scale	-4.258	10588	< 1.0 e-300 ***	-0.4004	419	3.987 e-93 ***

Visual features. The descriptive results in 1 indicate user preference towards choosing links on the left side of the screen and specifically links that are at the beginning of a Wikipedia article. The model results in 2 confirm these observations statistically. The positional lead feature has a clear positive coefficient in both the binomial and zero-truncated negative binomial model indicating that if a link is in the lead it is more frequently transitioned. Additionally, we can see that the right-body feature (links in the right 90% body region) indicates negative coefficients while the left-body feature (links in the left 10% body region) has a positive coefficient for the binomial model. The models for the infobox are contradicting each other, but overall show much less model improvements over the baseline model compared to our other visual feature models (lower LRT). What we can furthermore clearly see is that the navbox feature has a strong negative coefficient suggesting that links positioned at the bottom of the page in the navigation boxes are seldom visited. The coordinate features summarize the more fine-grained results in a broader level indicating that there is a preference towards choosing links that are at the beginning and also on the left side of the screen. The results presented are by and large confirming previous findings on the F-shaped way of information consumption [9, 13, 29].

Summary. Our results suggest three indicators having an effect on the success of links on Wikipedia. First, humans appear to have a *preference towards choosing links leading to the periphery* of the underlying topological link network. Second, *links connecting semantically related articles tend to be clicked more frequently* than those connecting semantically unrelated articles. Third, the visual position of a link appears to have an impact of the link’s success. Links that are located on the top of the screen (i.e., lead of an article) and on the left of the screen are preferred by humans navigating Wikipedia suggesting a *visual top-left preference*.

5. INTEGRATION IN MARKOV MODELS

While our results of Section 4 provide evidence of factors affecting the success of links on Wikipedia, it is an open question how we can integrate these findings into existing models of human

navigation. To that end, in this section, we investigate the influence of link features on models of human navigation. We focus on first-order Markov chains as the most wide-spread models for this task [5, 6, 37]. First order Markov chains are stochastic systems that assume a memoryless generative process, i.e., the probability of the next state (Wikipedia article) depends only on the current state. We utilize these models in two separate analyses. First, in Section 5.1, we craft hypotheses about human navigation based on our insights of Section 4.2 and incorporate these into a Bayesian inference process. This allows us to (a) identify whether given hypotheses can better explain observed transition behavior compared to a uniform structural baseline hypothesis and thus, potentially improve the Markov chain model fit, and (b) the model comparison provides an additional ranking of given hypotheses helping in confirming our previous results within a coherent research approach. Second, in Section 5.2, we aim at utilizing these insights to improve the classic PageRank algorithm in a weighted variation and validate its precision against the empirical visit rates of Wikipedia articles.

5.1 Bayesian integration

Next, we explore the impact of integrating our findings in Markov chain navigation models; for that task we resort to a recently proposed Bayesian approach called HypTrails [36].

Methodology. We aim to integrate hypotheses about navigational behavior, which we form based on our results obtained in Section 4.2, as prior knowledge into the inference process for Markov chains. HypTrails [36] provides a coherent framework for this task. First, one can express a hypothesis, i.e., beliefs in transition probabilities of a Markov chain, as a matrix $M = (m_{ij})$, where a higher value expresses higher belief in a transition between articles i to j . Then, these hypotheses matrices are incorporated as Dirichlet priors into the Bayesian model inference procedure. Due to the sensitivity of a Bayesian model to the priors on its parameters, more plausible hypotheses—i.e., hypotheses that are in line with the observed data—induce higher marginal likelihood (evidence) values used for model comparison. Marginal likelihoods can be compared along different

values of a parameter κ that expresses the strength of belief in a hypothesis. Higher values reflect a higher belief in given hypothesis; i.e., a stronger probability mass concentration of the Dirichlet prior. Formally, model comparison is done by calculating a *Bayes factor* [23], which is the ratio of the marginal likelihoods for two hypotheses H_1 and H_2 ; if positive, the first hypothesis is judged as more plausible—strength of the Bayes factor can be checked in an interpretation table provided by Kass and Raftery [23].

Comparing the marginal likelihood of a specific hypothesis (e.g., that humans prefer to navigate to the periphery of the network) with a structural baseline hypothesis (i.e., random preference of out-links) allows us to detect whether integrating given hypothesis into the inference process enhances the general model fit. Additionally, we can compare hypotheses with each other to obtain a ranking of hypotheses. For more methodological details, we point the reader to [36]. We describe how we construct hypotheses for our data next.

Hypotheses. We formalize and compare several hypotheses based on our main findings of Section 4 providing evidence of feature effects on link success. As mentioned, hypotheses are expressed as matrices $M = (m_{ij})$ indicating belief in transition probabilities. For finding good formalizations of these hypotheses, we experimented with different scaling variations such as logarithmic, square-root, and exponential feature scaling, but only report on the best formalization for each of the three feature groups due to limited space.

As a baseline, we use the *structural hypothesis* expressing the belief that someone navigating in the Wikipedia network chooses any out-link of an article randomly. Here, we set $m_{ij}^{structure} = 1$ if there is a link from article i to article j and $m_{ij}^{structure} = 0$ otherwise. If other hypotheses do not reveal higher marginal likelihoods than this baseline hypotheses (i.e., positive Bayes factors), they indicate no improvements to the model fit—i.e., they do not capture the actual human navigation behavior well.

For the network features, we focus on the *k-core* having shown promising results in Section 4. Since we want to formulate a hypothesis that expresses higher beliefs in transitions to articles with lower k-cores (i.e., the periphery of the network), we set the entries of the hypothesis matrix to $m_{ij}^{kcore} = \frac{1}{\sqrt{\text{trg_kcore}}}$. Regarding semantic similarity, we use the *text_sim* feature as it is, i.e., $m_{ij}^{text_sim} = \text{text_sim}$. This reflects a belief that transitions to similar articles (w.r.t. to the text similarity) are more likely. For visual features, we concentrate on the *position* feature. Our visual hypothesis states that links that are in the lead, in the most left part of the body, or in the infobox are more likely to be visited than the others (as the results of Section 4 suggest). Thus, we express this hypothesis as $m_{ij}^{visual} = 1$ in these cases, and to $m_{ij}^{visual} = 0$ otherwise. For these hypotheses, we add the structural baseline hypothesis for smoothing in order to guarantee a minimum belief in all transitions to linked articles.

In addition to these individual hypotheses, we also include combinations of them in our comparison. In this direction, we form the hypotheses *kcore+text_sim*, *kcore+visual*, *text_sim+visual*, and *kcore+text_sim+visual* by adding the respective matrices element-wise to each other—here, no smoothing is necessary.

Results. We report the resulting Bayes factors for the comparison between expressed hypotheses and the structural baseline hypothesis (for varying hypotheses weighting factors) in Figure 5; for a detailed description on how to read given figure, please resort to the figure caption. The primary observation is that all investigated hypotheses improve the fit of the Markov chain model in comparison to the structural baseline hypothesis as imminent from the Bayes factors for all values of κ . Complementary hypotheses (e.g., navigational preference towards the core of the network) result in negative Bayes factors falling below the baseline (not shown here). The relative

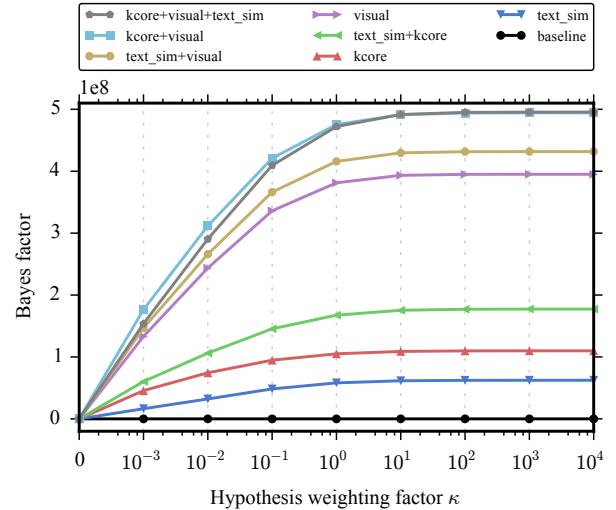


Figure 5: Bayesian integration results. This figure reports the results for our experiments on integrating navigational hypotheses as Bayesian priors in Markov chain models. The x-axis depicts different hypothesis weighting factors κ —higher values reflect stronger belief in a given hypothesis. The y-axis denotes the Bayes factor for a given hypothesis (different lines); the Bayes factor compares the fit of the respective hypothesis model to the fit of the baseline, i.e., the structural hypothesis. Higher Bayes factors reflect higher plausibility of given hypothesis; all Bayes factors in this plot are strongly decisive. All hypotheses improve the fit of the Markov chain model when integrated as prior assumptions. Hypotheses labels are in order of their ranking. A combination of visual and network features leads to the strongest improvement.

ranking of hypotheses also provides interesting insights. Comparing the single hypotheses (*text_sim*, *kcore* and *visual*) reveals that the visual hypothesis appears to be the most relevant, followed by the *kcore* and then the similarity hypothesis. Additionally, by combining these hypotheses, we can further improve the model. Overall, a combination of all three hypotheses has the highest evidence for higher values of κ . Yet, the combination of belief of peripheral navigation and preference towards choosing links in the visual top and left position already provides a similarly large improvement.

Summary. By and large, the results of this section have successfully demonstrated that we can improve models of human navigation on Wikipedia—i.e., the Markov chain model—by integrating hypotheses (beliefs) about human navigational behavior into the inference process of the models. In this case, we have utilized Bayesian inference to incorporate expressed hypotheses based on our insights of Section 4 as priors. All hypotheses improve the model fit. Additionally, a ranking of hypotheses reveals the relative plausibility and combinations show further improvements. Next, we want to further make use of these insights to improve the well-known PageRank algorithm utilizing the Markovian framework.

5.2 PageRank integration

In this section, we investigate if our insights can be used for improving the well-known PageRank algorithm for Wikipedia articles.

Weighted PageRank. The PageRank algorithm [30] is one of the most popular Web algorithms. It computes a centrality measure—*pagerank*—for each node in the network, such that a node receives a high value if it has many in-coming links from other important nodes. This can be interpreted as the probability of a random surfer in the network to land on respective page—i.e., the stationary distribution

of a Markov chain model. While for the classic PageRank algorithm, weights in the network are propagated equally through all links, we propagate weights according to the eight hypotheses that we formalized in Section 5.1 based on our main findings. That is, we model the landing probabilities of a random surfer that chooses the next node in the network proportional to the entries in the respective hypothesis matrix $M = (m_{ij})$. This weighted version of PageRank implements a *reasonable surfer model* [11], i.e., a model where a random surfer does not choose links uniformly, but is influenced by link features such as the link position. Formally, we compute the weighted pagerank of a node as:

$$PR(v) = \frac{1 - \alpha}{N} + \alpha \sum_{j \in \Gamma^-(i)} \frac{PR(i) \cdot m_{ij}}{Z}$$

where N is the number of nodes in the networks, $\Gamma^-(i)$ is the set of nodes linking to the node with index i , α is the damping factor and $Z = \sum_{j' \in \Gamma^-(i)} m_{ij'}$ is a normalization factor for each matrix row.

For evaluation, we determine the Spearman rank correlation of respective weighted PageRank with the sum of incoming transitions for a page from D_{trans_v} , i.e., number of views of an article stemming from internal navigation. As a baseline, we use the classic unweighted PageRank algorithm assuming random navigation.

Results. Table 3 shows the resulting Spearman correlation coefficient for the different evaluated hypotheses and the unweighted PageRank as a baseline for different damping factors α . We observe that most, but not all hypotheses achieve an improvement of the correlation between the aggregated transitions per target article and the pagerank values. In particular, the *kcore* and *visual* hypotheses lead to improvements. The best result can be obtained by combining these two hypotheses: the *kcore_visual* hypothesis achieves the best correlation score, which is about 0.1 better than the baseline. In contrast, and unexpected considering the previous results, the *text_sim* hypothesis is not able to increase the correlation, and also does not lead to improvements when combined with other hypotheses. This might be caused by the fact that PageRank assumes an infinite random walk while people navigate in line with this hypothesis maybe only in certain phases of their Wikipedia visit, e.g., at the end of their visit, cf. [42] and Section 7.

Apart from that, results are mostly in line with the ranking that we obtained from the Bayesian integration of the hypothesis, cf. Section 5.1. All correlation coefficients reported in this paper are strongly significant according to a t-test with a null hypothesis that two datasets are uncorrelated. Additionally, we formally test whether the improvements of the weighted PageRank algorithms in terms of their correlation coefficients with view statistics are significant compared to the unweighted algorithm. To that end, we employ Steiger’s one-tailed hypothesis test for assessing the difference between two paired correlations [39]; the results reveal clear significance for all correlations when compared to the baseline.

Summary. The results presented in this section suggest that a PageRank model with an intelligent edge weighting reflecting the users’ transition behavior explains the page views stemming from internal navigation better than the standard random surfer model. The reasonable surfer implemented through the edge weighting that performs best tend to select links that are leading *out of the network core to articles in the periphery and located in the top of the screen, on the left side of the screen and in the infoboxes*.

6. RELATED WORK

Since the inception of the Web, our research community has been interested in studying human navigational click data on the Web—e.g., see [19]. In this line of research, a variety of models have been

Table 3: **Weighted PageRank results.** Comparison of the hypotheses specified in Section 5.1 with respect to the Spearman correlation between the transitions (aggregated per target article) and the pagerank values of the corresponding articles in D_{wiki} for different damping factors. The correlation coefficient for the unweighted PageRank (marked *italic*) is used as a baseline. We see that most, but not all hypotheses achieve an improvement (marked **bold**); the best correlation with an improvement of ~ 0.1 across the three dumping factors is achieved by the hypothesis *kcore_visual*.

Damping factor α /Hypothesis M	0.80	0.85	0.90
<i>baseline</i>	<i>0.421</i>	<i>0.428</i>	<i>0.436</i>
<i>kcore</i>	0.434	0.440	0.447
<i>visual</i>	0.507	0.516	0.526
<i>text_sim</i>	0.400	0.407	0.415
<i>text_sim+kcore</i>	0.407	0.412	0.417
<i>text_sim+visual</i>	0.489	0.500	0.513
<i>kcore+visual</i>	0.530	0.538	0.545
<i>kcore+visual+text_sim</i>	0.494	0.505	0.517

proposed such as the well-known Markov chain model also utilized in this work [30, 34, 36, 37], or models such as decentralized search [12, 16] motivated by small-world navigation [24].

Insights have been utilized to infer missing links [43], predict the break-up of navigation process [35], for recommendations [41], or to improve the link structure of a website [31]. For the latter, Paranjape et al. [31] highlight the importance of improving hyperlink structures based on their usage due to the large amount of unused links. Consequently, they propose an algorithm for suggesting useful links and estimate their success based on clickthrough rates.

Based on this wide range of studies aiming at understanding human Web navigation, a series of navigational regularities, patterns and strategies have been suggested. For example, West and Leskovec [44] found trade-offs between similarity and degree in navigational behavior suggesting different phases in user sessions, namely an exploration (orientation) and an exploitation (goal-seeking) phase [16]. Subsequent research has suggested that humans prefer to navigate between semantically similar websites [36, 38, 42], have preferences for choosing links at the beginning of pages [13, 26, 31], and that navigational patterns exhibit regularities with respect to underlying network characteristics [26, 31, 36, 42].

While, as elaborated, a large amount of studies on human navigational behavior exist and a variety of navigational hypotheses have been proposed, there still has been a lack of a systematic understanding of what makes a link successful in information networks. To that end, we have extended previous work covered in this section by providing a study on effects of link properties on actual user transition behavior on the complete English Wikipedia utilizing large-scale data. Instead of providing a global view on clickthrough rates, we have also focused our attention on the clear setting of given a current Wikipedia article, which features best predict the popularity of links on that article. The features of interest have been motivated by previous work and our results broadly confirm previous hypotheses. Also, our analyses allow for more fine-grained insights such as navigational preference towards peripheral nodes or visual preferences on detailed screen coordinates. We have also demonstrated the importance of better understanding human navigational behavior and link popularity by successfully enriching existing Markov chain models and the PageRank algorithm with supplementary behavioral hypotheses.

Our work is also broadly connected to research studying click data in other contexts, e.g., characterizing user behavior and sybil detection in online social networks [4, 40], improving search en-

gine ranking functions [1, 20, 21, 22, 46], and marketing and next purchase prediction [7, 8, 10, 28]. However, these do not cover the specifics of large-scale information networks such as Wikipedia.

7. DISCUSSION

In this section, we discuss limitations and future work.

No session information. In this paper, we have focused on aggregated click counts for link transitions. While this was perfectly suited for our task at hand, we cannot differentiate between potentially varying phases within navigation sessions as e.g., zoom-in and zoom-out phases postulated in [42]. Also, we have only looked at *internal navigation* while observed behavior might differ when people navigate between different platforms (e.g., Google to Wikipedia).

Multiple link occurrence. Although the Wikipedia editor guidelines discourage placing multiple occurrences of links to the same target article within one source article, this is sometimes ignored by editors. Additionally, links can be repeated in the infoboxes or overview tables for improved usability. Unfortunately, the data at hand does not disambiguate which link instance has been clicked if a link occurs multiple times. While this has no implications for the network and semantic features as they are independent of the link position, it influences the analysis of the visual features. In this paper, we have decided to use the visual feature of the first occurrence of a link on the screen based on its x- and y-coordinates. This choice is justified by additional experiments on a restricted dataset filtering links with multiple occurrences. As for the main data, a strong preference for links in the top and the left hand side of the page can be found, cf. also Section 4.2. The results for the restricted dataset confirm our main findings, but also reveal a bias: the popularity of links at the top of the page is less pronounced (but still substantial) in this variation. For further analysis, we included the number of times a link occurs on an article as a nuisance parameter to our regression model, but this did not change the main inference. Thus, we are confident that our main findings are robust with respect to that issue. However, in future—preferably supplemented with more fine-grained click data—further studies on the impact of multiple link occurrences should be conducted.

Different screen resolutions. In our analysis, we chose the specific screen resolution of 1920×1080 for deriving visual features. While this adheres to our data only containing click information from desktop users only, different users may have different visual perception of the same article. Since this may affect presented results, future work could set out to tackle this limitation by, e.g., repeating the visual analysis for different screen resolutions. Another issue of special interest would be to complement the analysis using transitions data capturing the mobile user behavior on Wikipedia.

Definition of success. In this work, we defined success of a link as the click popularity derived from user transitions. However, other forms of link success can be studied. For example, instead of explaining transition preference of out-going links, one could modify the research question of this study and look at in-going links. Then, the question of interest would be: Given a set of in-going links to a given article, which ones are more popular than others and how can we explain them? Also, in this work, we have not considered the narrow textual context of a link which is tightly related to the concept of information scent [33]. Further research can concentrate on this aspect and study the relationship between diminishing information scent [15] caused, i.e., by editors and the success of a link. Apart from that, the success of a link could also be defined in light of how much a link contributes to the different types of navigation. One could analyze navigability of the networks D_{trans} and D_{trans_w} on the edge and node level, i.e., by studying the

flow in and out of a node or by removing nodes and edges. Link success could be also examined in terms of its ability to enable and encourage further navigation and exposing the user to a more richer content or to the strongly connected component of the network.

Markov chain hypotheses formulation. The results of Section 5 depend on how we code the hypotheses as priors. In this paper, we hand-crafted hypotheses that are in line by our data-driven insights of Section 4. We acknowledge that more fine-grained engineering and learning of good priors and weightings might further improve results. Additionally, even though in different context and with different methodology, one might argue that some results of Section 5 could have been expected given our earlier findings. Yet, the main goal here was to demonstrate the utility of incorporating assumptions about transition behavior into existing navigational models (i.e., Markov chain and PageRank) that predominantly, often have assumed uniform behavior. Additionally, obtained results helped us to further confirm our empirical insights of this paper. Nonetheless, we hope that in the future more studies on additional datasets, e.g., on other language editions of Wikipedia, are conducted in order to enrich our understanding of human navigation and link success.

Explaining and modeling heterogeneity. While we have accounted for basic heterogeneity by utilizing mixed-effects models and Markov chain models that both model transitions given a current start article, more fine-grained analyses in that direction are warranted in future. For example, there might be specific user groups, positions within sessions, or specific times that exhibit deviating navigation behavior. In this direction, a detailed analysis of the random effects in the mixed-effects models might lead to better understanding of heterogeneous effects. Additionally, pattern mining approaches could be employed to detect interpretable subgroups of Markovian transition behavior [27]. Obtained insights could then eventually be used to further advance existing models by integrating mixtures of navigational hypotheses into the modeling process.

8. CONCLUSIONS

In this paper, we have studied what makes a link successful on Wikipedia. To that end, (i) we investigated link features and their effects on link popularity utilizing descriptive analyses supplemented by mixed-effects hurdle models. Results suggest that Wikipedia users prefer to navigate to articles that are in the periphery of the Wikipedia link network, to articles that are semantically similar, and to articles that are linked at the top or at the most left-hand side of the source article. (ii) Based on these findings, we integrated hypotheses about human navigational behavior into the well-known Markov chain model utilizing Bayesian inference. By doing so, we could improve respective model fits compared to a uniform baseline model. (iii) For further demonstrating the utility of our findings, we adapted the well-known PageRank algorithm that assumes random navigation by accounting for observed navigational preferences in a weighted variation. An evaluation of resulting pagerank values obtained by link weighting against actual view statistics revealed significant improvements over the unweighted baseline algorithm.

Our work is relevant for researchers interested in studying human navigational behavior and link structures on the Web. Obtained insights can be utilized for providing qualitative link structures and for supplementing existing algorithms utilizing these structures to better adhere to observed transition behavior. Apart from further future work discussed in Section 7, we aim at extending our studies to different language versions of Wikipedia and newer timeframes for identifying similarities and differences. Also, while we have focused on Wikipedia in this work, the provided methodological framework can be applied to other platform data in a straight-forward manner.

References

- [1] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *International Conference on Research and Development in Information Retrieval*, 2006.
- [2] G. D. Bader and C. W. Hogue. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, 4(1):1, 2003.
- [3] D. Bates, M. Maechler, B. Bolker, S. Walker, et al. lme4: Linear mixed-effects models using eigen and s4. *R Package Version*, 1(7), 2014.
- [4] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida. Characterizing user behavior in online social networks. In *Internet Measurement Conference*, 2009.
- [5] J. Borges and M. Levene. Data mining of user navigation patterns. In *Web usage analysis and user profiling*, pages 92–112. Springer, 2000.
- [6] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine.[en línea]. *Disponible en Web*, 1998.
- [7] R. E. Bucklin and C. Sismeiro. A model of web site browsing behavior estimated on clickstream data. *Journal of Marketing Research*, 40(3):249–267, 2003.
- [8] R. E. Bucklin and C. Sismeiro. Click here for internet insight: Advances in clickstream data analysis in marketing. *Journal of Interactive Marketing*, 23(1):35–48, 2009.
- [9] G. Buscher, E. Cutrell, and M. R. Morris. What do you see when you're surfing?: using eye tracking to predict salient regions of web pages. In *Conference on Human Factors in Computing Systems*, 2009.
- [10] P. Chatterjee, D. L. Hoffman, and T. P. Novak. Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Science*, 22(4):520–541, 2003.
- [11] J. A. Dean, C. Anderson, and A. Battle. Ranking documents based on user behavior and/or feature data, 2010. US Patent 7,716,225.
- [12] D. Dimitrov, P. Singer, D. Helic, and M. Strohmaier. The role of structural information for designing navigational user interfaces. In *Conference on Hypertext and Social Media*, 2015.
- [13] D. Dimitrov, P. Singer, F. Lemmerich, and M. Strohmaier. Visual positions of links and clicks on wikipedia. In *International Conference Companion on World Wide Web*, 2016.
- [14] X. A. Harrison. Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ*, 2:e616, 2014.
- [15] M. Hearst. *Search user interfaces*. Cambridge University Press, 2009.
- [16] D. Helic, M. Strohmaier, M. Granitzer, and R. Scherer. Models of human navigation in information networks based on decentralized search. In *Conference on Hypertext and Social Media*, 2013.
- [17] J. J. Hox, M. Moerbeek, and R. van de Schoot. *Multilevel analysis: Techniques and applications*. Routledge, 2010.
- [18] M.-C. Hu, M. Pavlicova, and E. V. Nunes. Zero-inflated and hurdle models of count data with extra zeros: examples from an hiv-risk reduction intervention trial. *The American Journal of Drug and Alcohol Abuse*, 37(5):367–375, 2011.
- [19] B. A. Huberman, P. L. Pirolli, J. E. Pitkow, and R. M. Lukose. Strong regularities in world wide web surfing. *Science*, 280(5360):95–97, 1998.
- [20] T. Joachims. Optimizing search engines using clickthrough data. In *International Conference on Knowledge Discovery and Data Mining*, 2002.
- [21] T. Joachims. Evaluating retrieval performance using clickthrough data. Technical report, Cornell University, 2003.
- [22] S. Jung, J. L. Herlocker, and J. Webster. Click data as implicit relevance feedback in web search. *Information Processing & Management*, 43(3):791–807, 2007.
- [23] R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.
- [24] J. M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845–845, 2000.
- [25] K. Kristensen, A. Nielsen, C. W. Berg, H. Skaug, and B. Bell. Tmb: automatic differentiation and laplace approximation. *arXiv Preprint arXiv:1509.00660*, 2015.
- [26] D. Lamprecht, K. Lerman, D. Helic, and M. Strohmaier. How the structure of wikipedia articles influences user navigation. *New Review of Hypermedia and Multimedia*, pages 1–22, 2016.
- [27] F. Lemmerich, M. Becker, P. Singer, D. Helic, A. Hotho, and M. Strohmaier. Mining subgroups with exceptional transition behavior. In *International Conference on Knowledge Discovery and Data Mining*, 2016.
- [28] A. L. Montgomery, S. Li, K. Srinivasan, and J. C. Liechty. Modeling online browsing and path analysis using clickstream data. *Marketing Science*, 23(4):579–595, 2004.
- [29] J. Nielsen. F-shaped pattern for reading web content. <https://www.nngroup.com/articles/f-shaped-pattern-reading-web-content>. Accessed: 2016-07-12.
- [30] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: bringing order to the web. Stanford InfoLab, 1999.
- [31] A. Paranjape, R. West, L. Zia, and J. Leskovec. Improving website hyperlink structure using server logs. In *International Conference on Web Search and Data Mining*, 2016.
- [32] J. Pinheiro and D. Bates. *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media, 2006.
- [33] P. Pirolli. *Information foraging theory: Adaptive interaction with information*. Oxford University Press, 2007.
- [34] P. L. Pirolli and J. E. Pitkow. Distributions of Surfers' Paths through the World Wide Web: Empirical Characterizations. *World Wide Web*, 2(1-2):29–45, 1999.
- [35] A. T. Scaria, R. M. Philip, R. West, and J. Leskovec. The last click: Why users give up information network navigation. In *International Conference on Web Search and Data Mining*, 2014.
- [36] P. Singer, D. Helic, A. Hotho, and M. Strohmaier. Hyptrails: A bayesian approach for comparing hypotheses about human trails on the web. In *International Conference on World Wide Web*, 2015.
- [37] P. Singer, D. Helic, B. Taraghi, and M. Strohmaier. Detecting memory and structure in human navigation patterns using markov chain models of varying order. *PloS One*, 9(7):e102070, 2014.
- [38] P. Singer, T. Niebler, M. Strohmaier, and A. Hotho. Computing semantic relatedness from human navigational paths on wikipedia. In *International Conference on World Wide Web*, 2013.
- [39] J. H. Steiger. Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2):245, 1980.
- [40] G. Wang, T. Konolige, C. Wilson, X. Wang, H. Zheng, and B. Y. Zhao. You are how you click: Clickstream analysis for sybil detection. In *USENIX Security Symposium*, volume 9, pages 1–008, 2013.
- [41] Y. Wang, W. Dai, and Y. Yuan. Website browsing aid: A navigation graph-based recommendation system. *Decision Support Systems*, 45(3):387–400, 2008.
- [42] R. West and J. Leskovec. Human wayfinding in information networks. In *International Conference on World Wide Web*, 2012.
- [43] R. West, A. Paranjape, and J. Leskovec. Mining missing hyperlinks from human navigation traces: A case study of wikipedia. In *International Conference on World Wide Web*, 2015.
- [44] R. West, J. Pineau, and D. Precup. Wikispeedia: An online game for inferring semantic distances between concepts. In *International Joint Conference on Artificial Intelligence*, 2009.
- [45] E. Wulczyn and D. Taraborelli. Wikipedia clickstream. figshare. doi:10.6084/m9.figshare.1305770. Accessed: 2015-12-13.
- [46] G.-R. Xue, H.-J. Zeng, Z. Chen, Y. Yu, W.-Y. Ma, W. Xi, and W. Fan. Optimizing web search using web click-through data. In *International Conference on Information and Knowledge Management*, 2004.